

Statistical Science  
2014, Vol. 29, No. 1, 36–41  
DOI: 10.1214/13-STS424  
© Institute of Mathematical Statistics, 2014

# From Science to Management: Using Bayesian Networks to Learn about *Lyngbya*

Sandra Johnson, Eva Abal, Kathleen Ahern and Grant Hamilton

**Abstract.** Toxic blooms of *Lyngbya majuscula* occur in coastal areas worldwide and have major ecological, health and economic consequences. The exact causes and combinations of factors which lead to these blooms are not clearly understood. *Lyngbya* experts and stakeholders are a particularly diverse group, including ecologists, scientists, state and local government representatives, community organisations, catchment industry groups and local fishermen. An integrated Bayesian network approach was developed to better understand and model this complex environmental problem, identify knowledge gaps, prioritise future research and evaluate management options.

**Key words and phrases:** Bayesian statistics, Bayesian networks, *Lyngbya*.

## 1. INTRODUCTION

One of the most common marine pests in waterways around the world is algae. Harmful algal blooms occur across the world and have a wide range of detrimental impacts (Hamilton, McVinish and Mengersen, 2009). For example, they can replace or degrade other algal species that act as fish breeding grounds, poison fish and mammal marine life through the production of toxins (Arthur et al., 2006; Arthur et al., 2008), adversely affect coastal economies through reduced tourism and fishing (Watkinson, O’Neil and Dennison, 2005), and affect human health through dermatitis (Osborne, Webb and Shaw, 2001; Osborne, Shaw and Webb, 2007), neural disorders and contamination

of other seafood such as shellfish (Pittman and Pittman, 2005). One of the most common forms of harmful algae is cyanobacteria, or blue-green algae, and one of the most common species of cyanobacteria in tropical and subtropical coastal areas worldwide is *Lyngbya majuscula* (Dennison et al., 1999b; Arquitt and Johnstone, 2004). *Lyngbya*, also known as mermaid’s hair, stinging limu or fireweed, appears to be increasing in both frequency and extent (Dennison et al., 1999a; Albert et al., 2005). These blooms are due to a complex system of biological and environmental factors, exacerbated by human activities (Watkinson, O’Neil and Dennison, 2005). Thus, while there is a wealth of scientific and social literature on different aspects of the *Lyngbya* problem, for example, the role that nutrients play in the initiation and extent of *Lyngbya* blooms, or the effect of industry practices in the catchment on the nutrients available for *Lyngbya* growth, effective management of *Lyngbya* requires a “whole-of-system” approach that comprehensively integrates the different scientific factors with the available management options (Johnson et al., 2010). There is also a need to understand the different factors that trigger the initiation of a bloom versus the sustained growth of the cyanobacteria bloom.

Bayesian models are natural vehicles for describing complex systems such as these (Johnson and Mengersen, 2012). Key attributes of Bayesian models

---

Sandra Johnson is Ph.D., School of Mathematical Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia (e-mail: [sandra.johnson@qut.edu.au](mailto:sandra.johnson@qut.edu.au)). Eva Abal is Associate Professor, Office of the Deputy Vice Chancellor-Research, University of Queensland, Brisbane, QLD 4072 Australia (e-mail: [e.abal@uq.edu.au](mailto:e.abal@uq.edu.au)). Kathleen Ahern is Ph.D., Department of Environment and Heritage Protection, 400 George St, Brisbane, QLD 4000, Australia (e-mail: [kattyahern@gmail.com](mailto:kattyahern@gmail.com)). Grant Hamilton is Ph.D., School of Earth, Environmental and Biological Sciences, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia (e-mail: [g.hamilton@qut.edu.au](mailto:g.hamilton@qut.edu.au)).

in this context include flexibility of the model structure, the ability to incorporate diverse sources of information through priors and the provision of probabilistic estimates that take appropriate account of uncertainty in the system (McCann, Marcot and Ellis, 2006; Jensen and Nielsen, 2007; Hamilton, McVinish and Mengersen, 2009). A Bayesian network (BN) is a graphical Bayesian model that uses conditional probabilities to encode the strength of the dependencies between any two variables (Pearl, 1985). Causal and evidential inferential reasoning may be performed by the BN, depending on the nature of the dependencies (Pearl, 1985). BNs are increasingly used to model complex systems (Bromley et al., 2005). Variables in the model are represented by nodes, and links between variables are represented by directed arrows. Each node is then ascribed a probability distribution conditional on its parent nodes. The information used to develop these distributions can be obtained from a variety of sources, including data relevant to the system, related experiments or observations, literature and expert judgement (McCann, Marcot and Ellis, 2006; Jensen and Nielsen, 2007). A common practice is to discretise the variables into a set of states, resulting in a series of conditional probability tables; hence, under the assumptions of directional separation (d-separation, so that the nodes are conditionally independent) and the Markov property (so that the probability distribution of a node depends only on its parents), the target response node is quantified as the product of the cascade of conditional probability tables in the network (Uusitalo, 2007). The quantified model can then be used to identify influential factors, perform scenario assessments, identify configurations of node states that lead to optimal response outcomes and so on. BNs can be expanded into object-oriented and dynamic networks (Jensen and Nielsen, 2007; Johnson et al., 2010); they can include extensions such as decision, cost and utility nodes (Jensen and Nielsen, 2007); and they can be linked to other BNs to create systems of systems models.

In this paper we describe an integrated Bayesian network (IBN) approach developed by our research team to address the problem of *Lyngbya* blooms in Deception Bay, Queensland, Australia. With its proximity to Brisbane, Australia's third largest city, Deception Bay is a popular tourist destination in the Moreton Bay region. The many waterways feeding from intensive and rural agricultural activities into the bay and its use for commercial and recreational fishing put pressure on

the marine environment and compound the issues resulting from a *Lyngbya majuscula* bloom (Dennison et al., 1999a). Our project was undertaken as part of the *Lyngbya* Management Strategy funded by the local and Queensland Government's Healthy Waterways Program. The project team comprised a *Lyngbya* science working group and a *Lyngbya* management working group, representing diverse scientific disciplines, industry groups, government agencies and community organisations. The IBN is now a living part of the Healthy Waterways Program and has been expanded beyond Moreton Bay.

## 2. AN INTEGRATED BAYESIAN NETWORK FOR *LYNGBYA*

The IBN approach that we developed involved a "science model" linked to a "management model". The components of the IBN are detailed below.

### 2.1 The Science Model

The science BN [depicted in Supplemental Figure 1 (Johnson et al., 2014)] comprised the target node, "Bloom Initiation", and 22 other nodes which were identified by the *Lyngbya* science working group as potentially playing a key role in the initiation of a *Lyngbya* bloom (Johnson et al., 2014). It was transformed into an object-oriented BN with subnetworks describing water (comprising nodes for past and present rain, groundwater and runoff), sea water (tide, turbidity and bottom current climate), air (wind and wind speed), light (surface light, light quality, quantity and climate) and nutrients (dissolved concentrations of iron, nitrogen, phosphorus and organics, particulates, sediments nutrient climate, point sources and available nutrient pool) (Johnson et al., 2010). The nodes of the science model were quantified using a range of information sources and models, including process and simulation models, Bayesian hindcasting models, expert elicitation, published and grey literature, and data obtained from monitoring sites, industry records, research projects and government agencies (Johnson et al., 2010).

### 2.2 Science Model Extensions, Alternatives and Sub-Models

The science object-oriented BN model was further extended to incorporate temporal trends through a dynamic Bayesian network comprising five time slices, one for each of the summer months November to March (Johnson and Mengersen, 2009). Lag effects of

rainwater and groundwater runoff were incorporated in the object-oriented BN, allowing information and influence from one month to flow through to the next (Johnson and Mengersen, 2009).

Additional BNs were also constructed to more fully evaluate the *Lyngbya* problem. These included separate BNs to model *Lyngbya* biomass, duration and decay (as opposed to initiation), and a BN to focus on the critical two month summer period in which most *Lyngbya* initiations occur (as opposed to annual averages of rainfall and temperature used in the original model).

A variety of other statistical models were used to quantify some of the nodes of the BN. For example, random forest models were created to predict benthic photosynthetically active radiation (Kehoe et al., 2012) and Bayesian regression models were developed using data obtained from the monitoring stations in the catchment (Hamilton, McVinish and Mengersen, 2009). The latter data set comprised *Lyngbya* occurrences for each month during January 2000 to May 2007, a total of 77 observations, and monthly averages of minimum and maximum air temperature (as proxies for water temperature), solar exposure and amount of sky not covered by cloud (as proxies for light), and total rainfall (as a proxy for nutrients available in the water column), measured over the same period. A Bayesian probit time series regression model was developed to predict the monthly probability of bloom based on a total of 17 covariates, comprising five main effects, five first-order autoregressive terms and seven selected interactions. Covariate selection was performed using a Bayesian reversible jump Markov chain Monte Carlo algorithm and Bayesian model averaging was used to obtain a final predictive model. Eight of the 890 models identified by the algorithm accounted for over 75% of the posterior model probability, and the model comprising a single term, average monthly minimum temperature, accounted for almost 50%.

### 2.3 The Management Model

The aim of the management network [Supplemental Figure 2 (Johnson et al., 2014)] was to facilitate evaluation of options available to government agencies, communities and industry groups that could potentially influence the delivery of nutrients to Deception Bay. Nutrient point sources, such as industries (e.g., aquaculture, poultry) and council facilities (e.g., waste water treatment plants), and diffuse sources, such as landuse (e.g., grazing land, forestry) and urban activities (e.g., stormwater), were geographically located in the catchment. Each of these sources was then quantified

with respect to the probability of high or low emissions of different types of nutrients under current, planned and best practice scenarios. While not a Bayesian network in the sense of propagating these probabilities, the network structure was a valuable vehicle for collating and displaying this information.

A GIS-based nutrient hazard map for the catchment was then developed for each unit of land in the catchment, based on the nutrient emissions of the sources, the soil pH and soil type at each source location, and distance of the sources to the nearest waterway (Pointon et al., 2008). This included a nutrient risk rating which was interpreted as the perceived risk that there will be “enough” of that nutrient to cause an increase in growth, extent and duration of a *Lyngbya* bloom.

### 2.4 Creating and Using the Integrated Bayesian Network

The science BN and the management network described above were integrated via a water catchment simulation model that was developed as part of the *Lyngbya* project. The IBN was conceived as a series of steps, whereby a management intervention is proposed, and the management model is used to inform about the expected nutrient discharge into the Deception Bay catchment. The catchment model simulates the movement of these nutrients to the *Lyngbya* site in the Bay, and the science network then integrates this nutrient information with the other factors in the BN to determine the probability of bloom initiation.

We briefly discuss here three ways in which the IBN was interrogated to learn about *Lyngbya majuscula* bloom initiation in Deception Bay. First, the science BN provided an overall probability of *Lyngbya* bloom initiation based on the BN structure and its inputs. For example, in a typical year, as defined by the *Lyngbya* management working group, the probability of a bloom was predicted to be 0.28. Based on the dynamic network, this probability was much higher in the months of November and December and fell slightly in March.

Second, the IBN informed about important factors affecting this probability. For example, based on the science network, the seven most influential factors were available nutrient pool (dissolved), bottom current climate, dissolved iron, dissolved phosphorus, light and temperature. Based on both networks, the comparative impact of different management land uses on the probability of a bloom could be computed: these probabilities were lowest for waste water treatment

plant (0.23) and grazing (0.27), and highest for waste disposal (0.63), aquaculture (0.63) and poultry (0.62).

Third, the IBN facilitated the evaluation of scenarios, for example, about the impacts of management options such as upgrading nominated point sources from current to best practice (e.g., eliminating potassium output from sewage treatment plants), climate events (e.g., a severe storm) and conditions most or least favourable for bloom initiation. For example, under optimal light climate and high temperature conditions, a storm event increased the probability of bloom initiation from 0.28 to 0.42 and initiation was certain if the available nutrient pool (dissolved) was enough. As another example, changing the management land use from natural vegetation to agriculture throughout the catchment area (based on the management network) results in an increase of 8.8% in available nutrients compared with baseline levels (based on the GIS hazard map), which in turn results in a substantial increase in the probability of a *Lyngbya* bloom initiation to 0.62 (based on the science BN). Note that the effect of this land use change is diluted by the fact that the proportion of the catchment designated as natural vegetation is only 18.24%.

Investigation of the BN also revealed unexpected results that required discussion and reflection by the science and management teams. For example, the model supported early suggestions that iron was a key nutrient in *Lyngbya* bloom initiation (Watkinson, O'Neil and Dennison, 2005), which motivated additional research into this important issue (Ahern, Ahern and Udy, 2008). As another example, land runoff and point sources contributed approximately equally to the probability of bloom initiation under the developed science model, provoking questions about the relative effects of population pressure and industrial growth in the catchment. Alternatively, it suggests that the information available to quantify these nodes is somewhat uncertain. In fact, it is a methodological challenge to accurately model the nutrient load into Deception Bay from land runoff (Kehoe et al., 2012) and more accurate models are currently under development.

### 3. WHY BAYESIAN?

By their nature, a complex system is challenging to model—using traditional statistical approaches. This is illustrated well in the *Lyngbya* case study described here, which is characterised by multiple interacting factors drawn from science and management, piecemeal knowledge and diverse information

sources (Kehoe et al., 2012). Furthermore, Bayesian models are able to capture the uncertainty in the data and parameter estimates which is generally agreed to be lacking in many ecological modelling paradigms (Hamilton, McVinish and Mengersen, 2009). More specifically, Bayesian networks (BNs) are capable of diagnostic, predictive and inter-causal (or “explaining away”) reasoning (Jensen and Nielsen, 2007; Johnson and Mengersen, 2012), which was particularly relevant for the *Lyngbya* problem described here.

There are several alternatives to the IBN approach that could be considered for modelling the *Lyngbya* problem. Janssens et al. (2006) proposed a decision tree approach, but this was less able to represent the many interactions between the factors in the system. Other methods include stochastic petri nets which are able to model concurrent systems (Angeli, De Leenheer and Sontag, 2007), but require the modeller to have advanced statistical knowledge and were unlikely to engage the diverse group of *Lyngbya* stakeholders. Process-based modelling, which is commonplace in ecology, requires substantial data for calibration and validation of the models, which is very time consuming and resource hungry and may take several years (Kehoe et al., 2012). In contrast, a BN allows us to assimilate current knowledge and modelling effort without having to wait until “perfect” and “sufficient” data are available. This is particularly important when dealing with a major environmental hazard such as toxic algal blooms. None of the alternative approaches had the unique combination of qualities of BNs which integrated the different sources of information, represented the dependencies and uncertainty in the information, guided future data collection and research, and engaged a diverse group of stakeholders.

The IBN described in this paper is the most comprehensive local systems model of *Lyngbya* that has been developed to date. There are many other examples of the use of BNs to solve “big” problems. We have employed them to investigate infection control in medicine, airport and train delays, wayfinding, import risk assessment (Mengersen and Whittle, 2011), peak electricity demand and sustainability of the dairy industry in Australia. Furthermore, there are other conceptual and methodological approaches to constructing BNs; examples include decision making in business (Baesens et al., 2004) and protein networks in biology (Jansen et al., 2003).

Finally, BNs are just one tool in the kit of statistical methods that should be considered for solving these



types of problems and that can be considered as complements to other approaches in order to reveal the full picture of a complex system.

### ACKNOWLEDGEMENT

The authors gratefully acknowledge Kerrie Mengersen for her significant contribution to this manuscript.

### SUPPLEMENTARY MATERIAL

**Supplementary Figures** (DOI: [10.1214/13-STS424SUPP](https://doi.org/10.1214/13-STS424SUPP); .pdf). Diagrams for the *Lyngbya* management network and the *Lyngbya* science Bayesian network are included in the supplemental article to this paper.

### REFERENCES

- AHERN, K. S., AHERN, C. R. and UDY, J. W. (2008). In situ field experiment shows *Lyngbya majuscula* (cyanobacterium) growth stimulated by added iron, phosphorus and nitrogen. *Harmful Algae* **7** 389–404.
- ALBERT, S., O'NEIL, J. M., UDY, J. W., AHERN, K. S., O'SULLIVAN, C. M. and DENNISON, W. C. (2005). Blooms of the cyanobacterium *Lyngbya majuscula* in coastal Queensland, Australia: Disparate sites, common factors. *Mar. Pollut. Bull.* **51** 428–437.
- ANGELI, D., DE LEENHEER, P. and SONTAG, E. D. (2007). A Petri net approach to the study of persistence in chemical reaction networks. *Math. Biosci.* **210** 598–618. [MR2374484](https://doi.org/10.1016/j.mbs.2007.03.004)
- ARQUITT, S. and JOHNSTONE, R. (2004). A scoping and consensus building model of a toxic blue-green algae bloom. *System Dynamics Review* **20** 179–198.
- ARTHUR, K. E., LIMPUS, C. J., ROELFSEMA, C. M., UDY, J. W. and SHAW, G. R. (2006). A bloom of *Lyngbya majuscula* in Shoalwater Bay, Queensland, Australia: An important feeding ground for the green turtle (*Chelonia mydas*). *Harmful Algae* **5** 251–265.
- ARTHUR, K., LIMPUS, C., BALAZS, G., CAPPER, A., UDY, J., SHAW, G., KEUPER-BENNETT, U. and BENNETT, P. (2008). The exposure of green turtles (*Chelonia mydas*) to tumour promoting compounds produced by the cyanobacterium *Lyngbya majuscula* and their potential role in the aetiology of fibropapillomatosis. *Harmful Algae* **7** 114–125.
- BAESENS, B., VERSTRAETEN, G., VAN DEN POEL, D., EGMONT-PETERSEN, M., VAN KENHOVE, P. and VAN THIENEN, J. (2004). Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European J. Oper. Res.* **156** 508–523.
- BROMLEY, J., JACKSON, N. A., CLYMER, O. J., GIACOMELLO, A. M. and JENSEN, F. V. (2005). The use of Hugin to develop Bayesian networks as an aid to integrated water resource planning. *Environmental Modelling Software* **20** 231–242.
- DENNISON, W., ABAL, E. G., ROGERS, J., COLLIER, C., GAUS, C. and SOUTH EAST QUEENSLAND REGIONAL WATER QUALITY MANAGEMENT STRATEGY (1999a). *Moreton Bay Study: A Scientific Basis for the Healthy Waterways Campaign*. South East Queensland Regional Water Quality Management Strategy, Brisbane, QLD.
- DENNISON, W. C., O'NEIL, J. M., DUFFY, E. J., OLIVER, P. E. and SHAW, G. R. (1999b). Blooms of the cyanobacterium *Lyngbya majuscula* in coastal waters of Queensland, Australia. *Bulletin de l'Institut Oceanographique, Monaco* **19** 501–506.
- HAMILTON, G., MCVINISH, R. and MENSERSEN, K. (2009). Bayesian model averaging for harmful algal bloom prediction. *Ecol. Appl.* **19** 1805–1814.
- JANSEN, R., YU, H., GREENBAUM, D., KLUGER, Y., KROGAN, N. J., CHUNG, S., EMILI, A., SNYDER, M., GREENBLATT, J. F. and GERSTEIN, M. (2003). A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* **302** 449–453.
- JANSSENS, D., WETS, G., BRIJS, T., VANHOOF, K., AR-ENTZE, T. and TIMMERMAN, H. (2006). Integrating Bayesian networks and decision trees in a sequential rule-based transportation model. *European J. Oper. Res.* **175** 16–34.
- JENSEN, F. V. and NIELSEN, T. D. (2007). *Bayesian Networks and Decision Graphs*, 2nd ed. Springer, New York. [MR2344166](https://doi.org/10.1007/978-1-4020-8829-5)
- JOHNSON, S. and MENSERSEN, K. (2009). A Bayesian network approach to modelling temporal behaviour of *Lyngbya majuscula* bloom initiation. In *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation* (R. S. Anderssen, R. D. Braddock and L. T. H. Newham, eds.). Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Cairns, Queensland, Australia.
- JOHNSON, S. and MENSERSEN, K. (2012). Integrated Bayesian network framework for modeling complex ecological issues. *Integr. Environ. Assess. Manag.* **8** 480–490.
- JOHNSON, S., FIELDING, F., HAMILTON, G. and MENSERSEN, K. (2010). An Integrated Bayesian Network approach to *Lyngbya majuscula* bloom initiation. *Mar. Environ. Res.* **69** 27–37.
- JOHNSON, S., ABAL, E., AHERN, K. and HAMILTON, G. (2014). Supplement to “From science to management: Using Bayesian networks to learn about *Lyngbya*.” DOI:[10.1214/13-STS424SUPP](https://doi.org/10.1214/13-STS424SUPP).
- KEHOE, M., O'BRIEN, K., GRINHAM, A., RISSIK, D., AHERN, K. S. and MAXWELL, P. (2012). Random forest algorithm yields accurate quantitative prediction models of benthic light at intertidal sites affected by toxic *Lyngbya majuscula* blooms. *Harmful Algae* **19** 46–52.
- MCCANN, R. K., MARCOT, B. G. and ELLIS, R. (2006). Bayesian belief networks: Applications in ecology and natural resource management. *Canadian Journal of Forest Research* **36** 3053.
- MENSERSEN, K. and WHITTLE, P. (2011). Improving accuracy and intelligibility of decisions. *Journal für Verbraucherschutz und Lebensmittelsicherheit* **6** 15–19.
- OSBORNE, N. J., SHAW, G. R. and WEBB, P. M. (2007). Health effects of recreational exposure to Moreton Bay, Australia waters during a *Lyngbya majuscula* bloom. *Environ. Int.* **33** 309–314.
- OSBORNE, N. J. T., WEBB, P. M. and SHAW, G. R. (2001). The toxins of *Lyngbya majuscula* and their human and ecological health effects. *Environ. Int.* **27** 381–392.

- PEARL, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. In *Seventh Annual Conference of the Cognitive Science Society, Irvine, CA*. Univ. California, Irvine.
- PITTMAN, S. J. and PITTMAN, K. M. (2005). Short-term consequences of a benthic cyanobacterial bloom (*Lyngbya majuscula* Gomont) for fish and penaeid prawns in Moreton Bay (Queensland, Australia). *Estuarine, Coastal and Shelf Science* **63** 619–632.
- POINTON, S. M., AHERN, K. S., AHERN, C. R., VOWLES, C. M., ELDERSHAW, V. J. and PREDA, M. (2008). Modelling land based nutrients relating to *Lyngbya majuscula* (Cyanobacteria) growth in Moreton Bay, southeast Queensland, Australia. In *Thirteenth International Marine Biological Workshop, The Marine Fauna and Flora of Moreton Bay, Queensland* 54 (P. J. F. Davie and J. A. Phillips, eds.) 377–390. Memoirs of the Queensland Museum Nature, Brisbane.
- UUSITALO, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling* **203** 312–318.
- WATKINSON, A. J., O'NEIL, J. M. and DENNISON, W. C. (2005). Ecophysiology of the marine cyanobacterium, *Lyngbya majuscula* (Oscillatoriaceae) in Moreton Bay, Australia. *Harmful Algae* **4** 697–715.